

Data Quality Report

1. Summary of the dataset

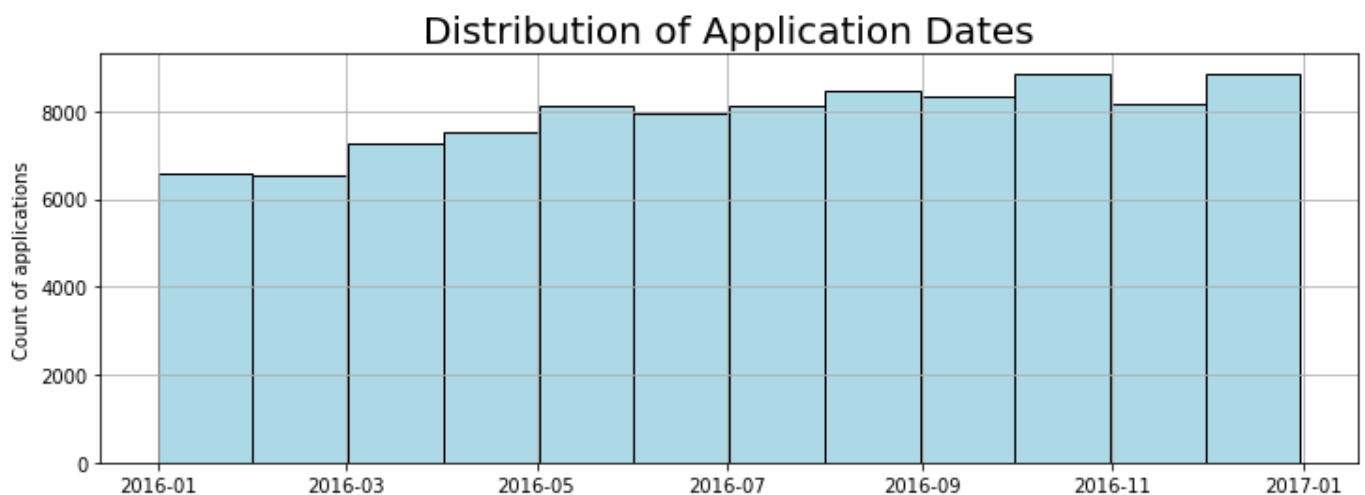
Dataset *applications* contains 10 variables and 94,866 observations. Transformations are made on fields with time stamps to reduce ambiguity.

	record	date	ssn	firstname	lastname	address	zip5	dob	homephone	fraud
count	94866	94866	94866	94866	94866	94866	94866	94866	94866	94866
unique	94866	365	86771	14626	31513	88167	15855	30599	20762	2
populated_rate	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

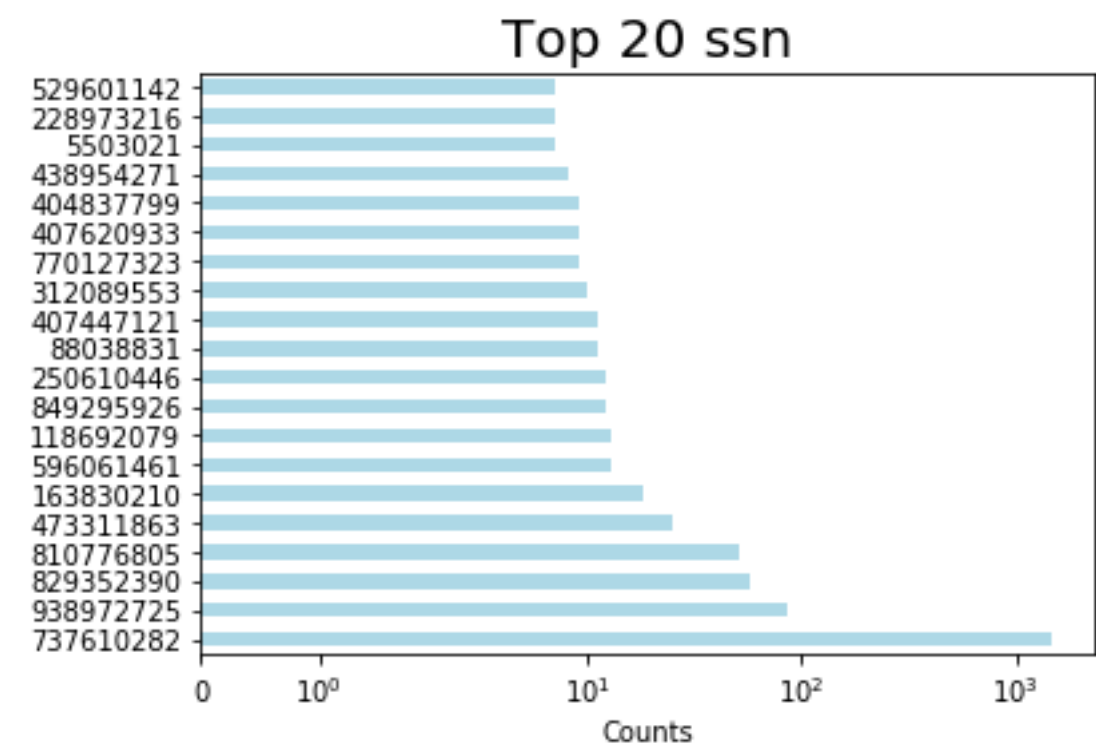
2. Detailed information for Each Field

Field Name	Description
record	Index of the dataset

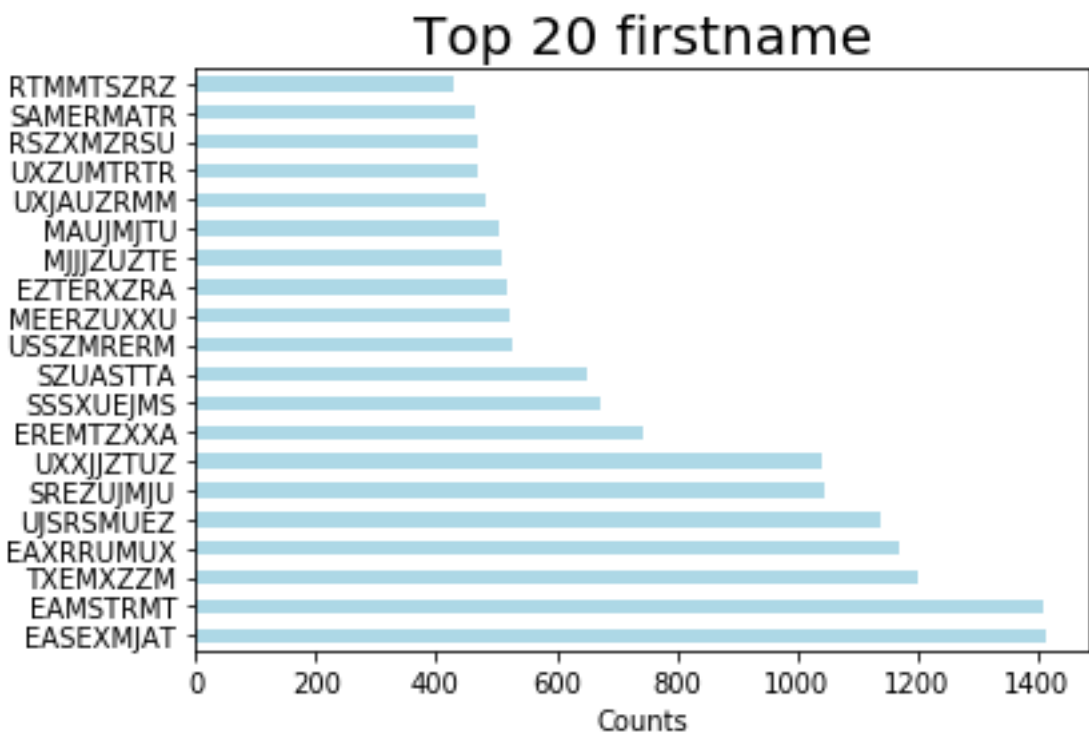
Field Name	Description
date	The application date



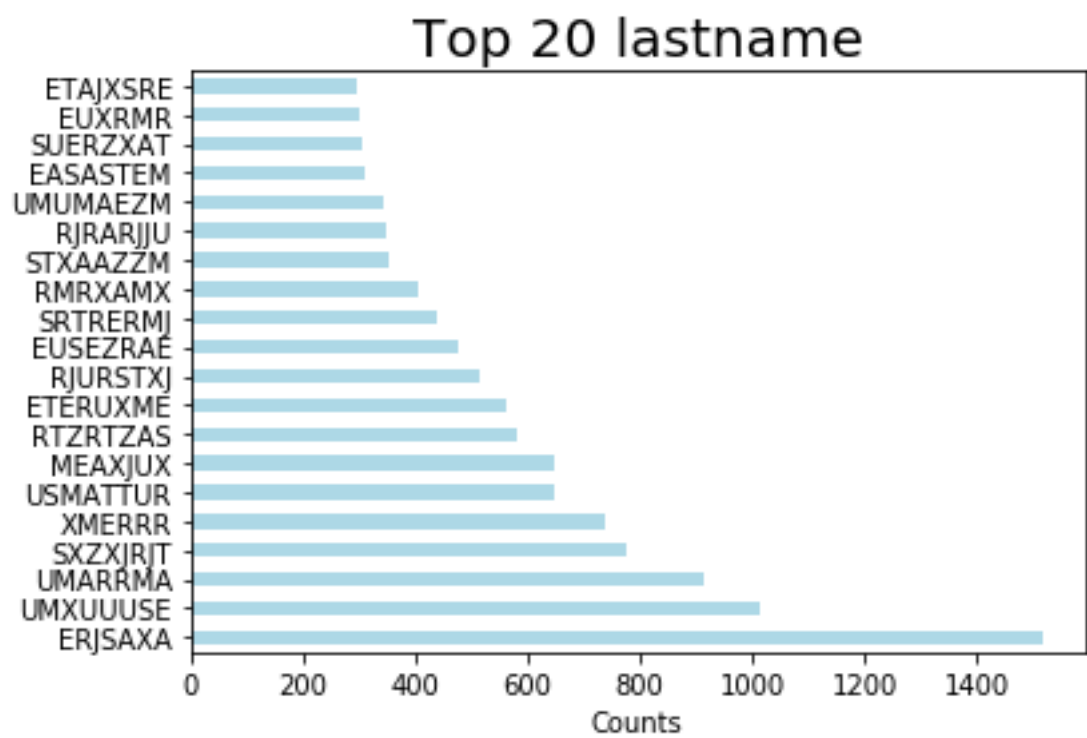
Field Name	Description
ssn	Social security number of the applicant



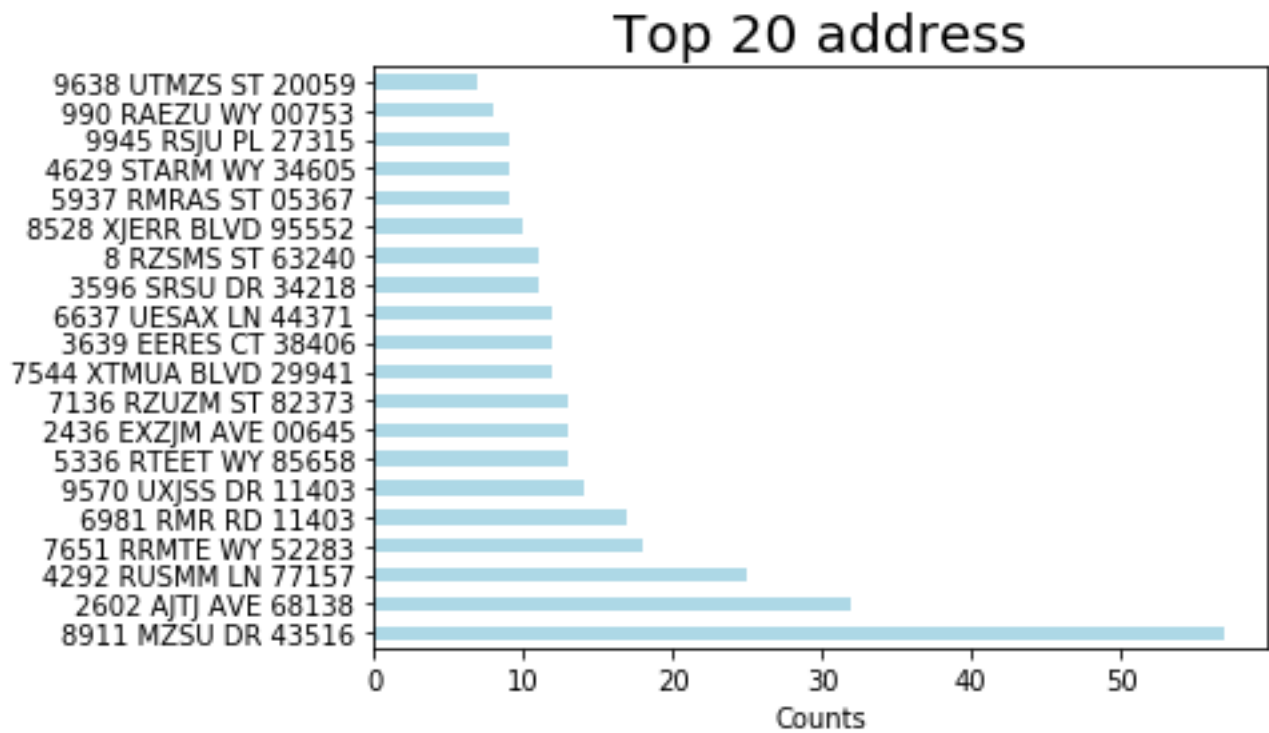
Field Name	Description
firstname	First name of the applicant



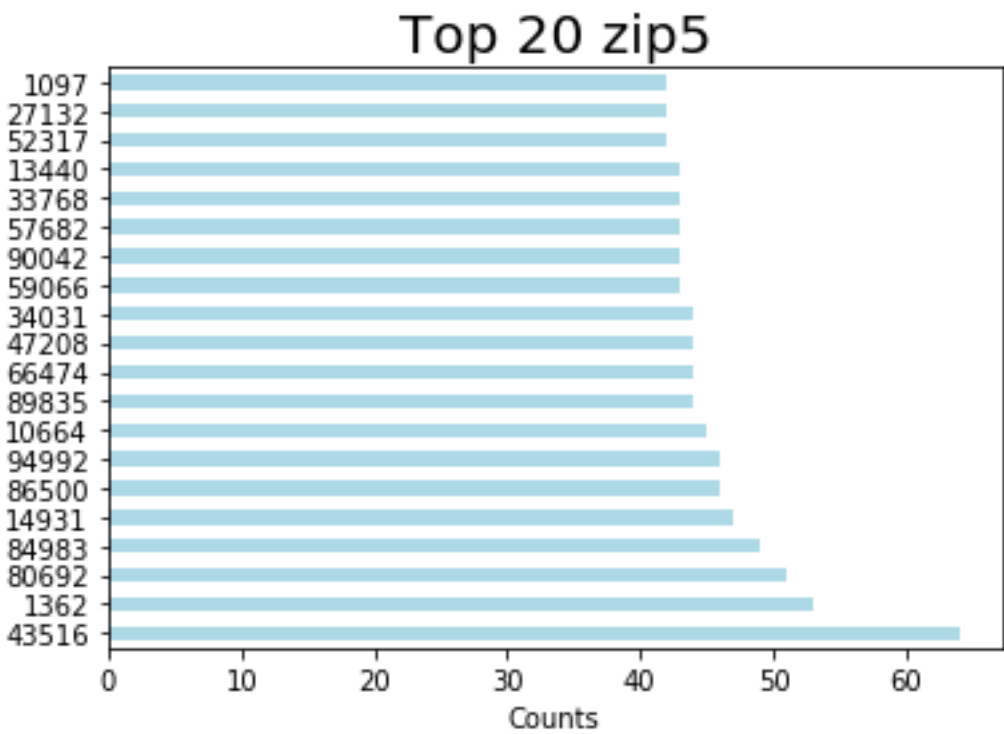
Field Name	Description
lastname	Last name of the applicant



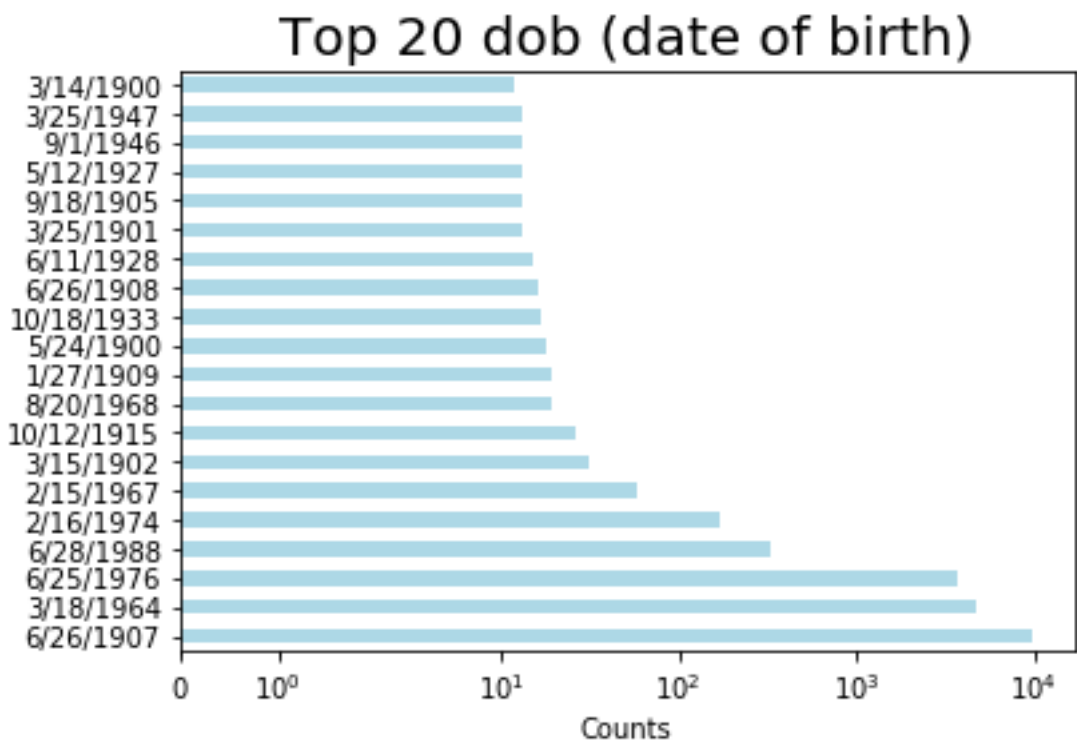
Field Name	Description
address	Address of the applicant



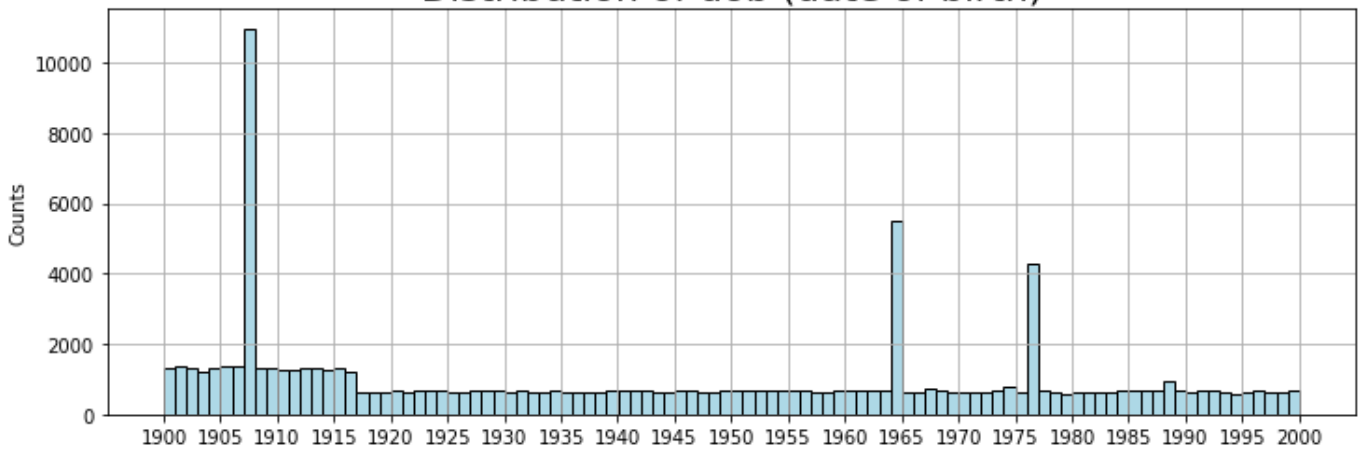
Field Name	Description
zip5	First five digits of zip code of the applicant



Field Name	Description
dob	Date of birth of the applicant

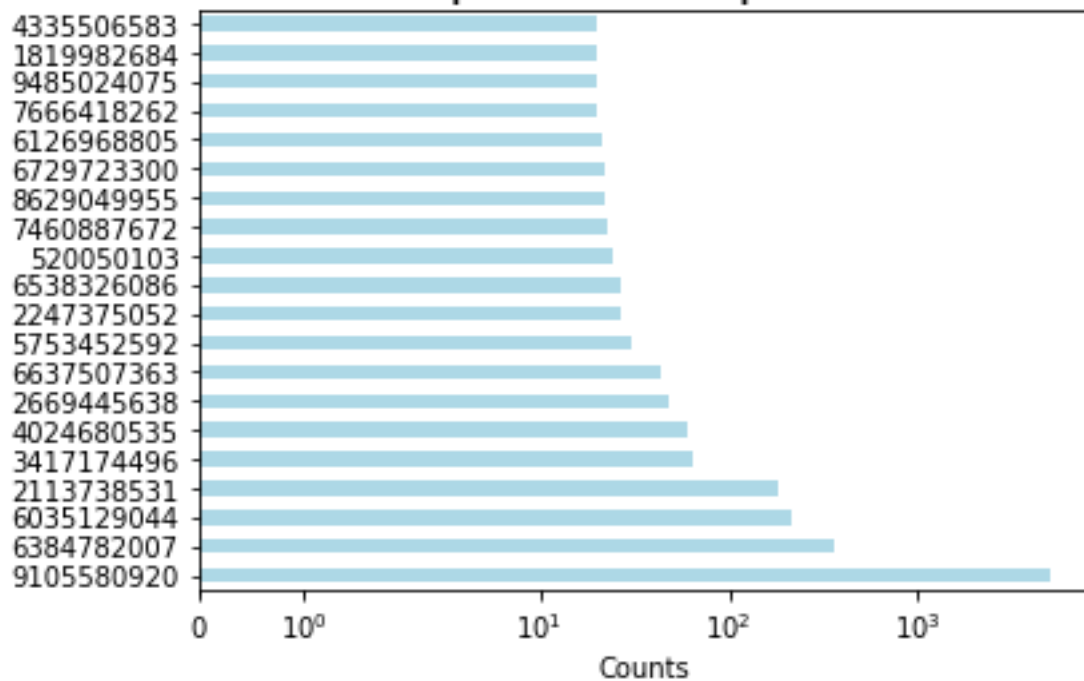


Distribution of dob (date of birth)



Field Name	Description
homephone	Phone number of the applicant

Top 20 homephone



Field Name	Description
fraud	Whether a certain record is a fraud or not.

